
Resource Guide: Voice Recognition & Real-time Technologies



A product of a State Justice Institute demonstration

Grant Number SJI-99-N-005 – Missouri Office of State Courts Administrator

Prepared by: Michael H. Zajdel , Court Reporter

Acknowledgements

Special thanks goes to the three pilot reporters for their hard work and dedication to this project. Many hours of personal time were devoted to their efforts in testing and evaluating this technology. The pilot reporter's judges are also owed special thanks for their patience and assistance in this process. They were:

29th Judicial Circuit

Sharon Lambert, Court Reporter

The Honorable David C. Dally

16th Judicial Circuit

James Bouck, Court Reporter

The Honorable Ronald R. Holliger

28th Judicial Circuit

Mike Zajdel, Court Reporter

The Honorable David Darnold

Disclaimer: This document was developed under grant number SJI-99-N-005 from the State Justice Institute. The points of view expressed are those of the author(s) and do not necessarily represent the official position or policies of the State Justice Institute.

Table of Contents

WHAT DO I NEED FOR SPEECH RECOGNITION	1
The Basics	1
HOW DO YOU INTEND TO WORK?	3
Redictation	4
Digital Recording and Transcription	5
Realtime	6
THE HARDWARE	8
Dictation System	8
Transcription System	8
Realtime System	9
THE SOFTWARE	11
Speech Recognition Software	11
CAT	12
AN AUDIO QUANDARY	13
TO DICTATE OR REDICTATE, THAT IS THE QUESTION	17

WHAT DO I NEED FOR SPEECH RECOGNITION

If you are either considering upgrading your existing system to run speech recognition software or are already using speech recognition software and are wanting to possibly increase your system's recognition accuracy, you have many choices to make, and there can be many variables affecting the logic of your decision. Unfortunately, there is no one easy solution for everyone's situation.

THE BASICS

The basic hardware components needed for speech recognition are:

- A PC
- Audio input device
- Speech recognition software

Before going further and elaborating on system requirements, you need to understand that just because your system meets the “minimum requirements” of what a software manufacturer suggests to run its hardware does not mean you will get optimal results. If your system barely meets the minimum requirements, it will give minimum results. When the software manufacturer claims dictation speeds up to 160 words per minute, it is based upon testing using state-of-the-art systems at the time the software was distributed, not a system that meets the minimum requirements.

Also, keep in mind that CPU speed is not the single determining factor in a system's performance. Not all processors are created equal, nor are system buses or subsystems. All the components taken together will determine how your system performs, so two 450 Mhz systems will not perform equally if they have different processors, say, an Intel Celeron and an Intel Pentium II. The Pentium II is a more robust chipset that has more memory and a faster bus, and accordingly it will always outperform a Celeron. That is why they are priced differently.

I have used my share of software products over the years and, excluding computer games and a teenager, I haven't seen any product push a

computer to its limit like speech recognition. No software product currently claims to be able to accurately recognize speech at over 160 words per minute. As court reporters, we push speech recognition technology to its limits, beyond what software manufacturers are currently claiming, to achieve good results at dictation speeds over 200 wpm. This is why system quality is so important to us.

Let me give you an example. I have a PII-450 system with 256 Megabytes of RAM, a high-quality sound card (Turtle Beach Fiji) and a very respectable 9 GB Ultra Wide SCSI hard drive. When I first got my system a year ago, matched with Dragon Naturally Speaking software that was written for the PII processor, I got what I considered to be very good results. A year after purchasing my system, there are now PIII and Athlon processors running at 800 Megahertz, and Dragon Systems has rewritten their software to capitalize on the increased computing power these processors offer. My once powerful system now seems inadequate.

If I use Dragon's newest feature, BestMatch III, when dictating around 180 wpm and higher, my system quickly falls behind, yielding worse results than I had using the old software. If, however, I use the new software's BestMatch III feature for merely transcribing, which allows the system to take its time to produce a record from a wave file, I see increased recognition over the previous version I was using. This leads me to the logical conclusion that I need at least a new CPU to get my system back up to par with the software if I want to use the software's newer, better features when doing realtime. My system is, for the purposes I want to use it for, inadequate, even though it is still way above the minimum requirements stated by the software manufacturer.

If you don't intend to do realtime, but only want to become familiar with speech recognition before investing in a high-end or complete CAT system, your system requirements are considerably less.

If there is not a realtime requirement, your system will not be under as great a burden, therefore lessening the hardware and software requirements. Keep in mind, however, that if your long-term goal is ultimately realtime, you will probably not want to skimp too much, because you will wind up needing newer, faster hardware down the road. That said, however, if you have a PC system that is borderline, and you aren't sure how well it will work, you might as well buy the software first and try it on that system. You really have nothing to lose, because if it doesn't work well enough on that system, you simply know it's time for an upgrade. Use your own common sense in this area.

HOW DO YOU INTEND TO WORK?

The first choice you need to make is to determine how you intend to use your system. Different working environments and job requirements create needs specific to the individual. The variations on implementing speech recognition (SR) can be somewhat confusing, so let's break it down a bit. Below are three ways court reporters can use SR.

- Redictation of proceedings.
- Creating digital recording on site, post-dictation transcribing.
- Realtime on site.

Realtime is obviously now a voice reporter's ultimate goal, but, for some, realtime is going to be more than they need at first, especially while speech recognition products are in their infancy, which as impressive as they seem to be, they are. For those who want or need realtime and all the capabilities that come with it, the technology is falling into place, but there are many factors that must be considered before doing so. When considering realtime, one should not take for granted the complex nature of these systems at the current time, because unrealistic expectations and inexperience can meet with disastrous results.

How you begin implementing SR and how you ultimately utilize it may change as technologies emerge and converge, and as your own situation changes. For many, the first factor to consider is how familiar you are with computers and SR products in general.

A novice in both these areas obviously has a steep learning curve ahead of him or her, and would probably do best developing proficiency incrementally in these areas before relying on SR technology in their daily work environment.

On the other hand, someone who has been working with PCs, feels comfortable in a Windows environment, and has been using speech recognition products for some time might easily make a jump to a realtime system. Many other factors come into play as well, so I'll try to discuss some of them now.

REDICTION

Redictating is probably the easiest and safest way to become familiar with speech recognition software. Redictation does not require any additional equipment on site, but it is not the most efficient method to produce a transcript. The reporter must first report the deposition or hearing and then must come home and dictate again. I would recommend redictation for the following users:

- Users who work in a fast-moving and/or hostile environment, such as certain courtroom environments.
- Users who are interested in speech recognition, have a desktop PC and some time to learn.
- Users who do not adapt to change well, especially things of a technical nature.
- Users who got their first PC this Christmas.

Hostile or frenetic environments are not conducive to making a good record in any situation, and speech recognition technology is no exception. Current SR systems do not give good results at consistently high rates of speech, no matter how well you dictate. At least, not yet. If you are routinely blue-faced during your normal day in court from lack of oxygen due to the speed of proceedings, you might as well not bother trying SR realtime in the courtroom at this point. Transcribing from a digital recording won't give satisfactory results, either; you'll be faster redictating, or even typing, on your own. I speak from experience regarding this matter.

For some, visibility might be an issue. Many reporters may wish to remain as invisible as possible during proceedings, or perhaps they work for some stuffy old judge who simply won't have any such nuisance as SR in his or her courtroom. Having a PC in front of you spitting out page after page of transcript as people speak definitely gets noticed nowadays, and it's not necessarily always a good thing.

In one of the four courtrooms I work out of, the witness stand is literally behind me, believe it or not. One day while using my system in that courtroom, a witness literally stopped talking in the middle of his answer. After a prolonged pause, I looked around to see what was the matter and saw that his eyes were glued to my screen. He was freaked out. After that incident, I always remember to reduce the font size on my screen so it can't be read from the witness chair, or to arrange my PC so it can't be viewed by anyone but me.

Transcripts can be created by redictation after proceedings have been taken, which is the invisible way to implement SR. The only person who sees the change is you, but the final result is identical.

For the user who is curious about the capabilities of SR, or who wants to develop their skills before committing to a full CAT system, it would probably be easiest and most economical to simply record and transcribe post-dictation. There is not much additional pressure on the reporter while taking proceedings, hardware cost is minimized, and the reporter can gradually develop his or her style of dictation to eventually facilitate software transcription and/or realtime by using SR software.

If you just don't feel comfortable with change, or are afraid of all things technical, but you know you are going to have to adapt to CAT down the road, the best and safest way to do it is incrementally. The first step for a person like you is to start by redictation. No one gets hurt this way.

DIGITAL RECORDING AND TRANSCRIPTION

This method gives a great deal of flexibility to the court reporter who doesn't need realtime on site but is willing to use a digital recording system for subsequent transcription via SR software. With this method, the reporter dictates the record using a digital recording device, then the audio file is later transcribed by a computer at the reporter's convenience. Assuming a high-quality digital recording has been made, a transcript with higher recognition accuracy than that of a realtime record can sometimes be made. I would recommend digital recording and transcription for the following users:

- Users who travel often.
- Users who work in courtroom environments.
- Realtime system owners.
- Users who are curious about SR and have a laptop PC.

Many will find this method easy to adapt to and very convenient. It can be used with a minimum of effort, can be a lightweight system, and is more convenient than redictation because the original dictation is used to create the transcript. This method also allows the reporter to minimize costs by utilizing pre-existing equipment.

Older model laptops that are not capable of realtime records often can make audio files good enough for software transcription. The same laptop that was used to record that audio may also be able to transcribe if it is a newer model. If the laptop is not powerful enough, the audio file will need to be transferred to another computer for transcription. Transfer can be

accomplished via CD media, high-capacity portable drive, cable, or whatever means is available.

Reporters using realtime systems who don't know the nature of the proceeding they are taking beforehand can subsequently train their SR user files post-dictation to get higher levels of accuracy than the original. Non realtime users can also train their SR user files post-dictation to increase recognition accuracy, only they won't have a realtime record, just transcription.

My realtime system can make a good realtime record, but it could not previously record an audio file and create a realtime record simultaneously. I had no demand for realtime in the court I work in, so I chose to record audio files during trials for software transcription. Subsequent comparison of unedited realtime transcripts and transcripts created via software transcription showed that the ones created post trial were more accurate, requiring less time to edit. Subsequent training of the system and retranscription of audio files in areas where recognition accuracy had dropped resulted in increased recognition accuracy.

REALTIME

This is it. The ultimate goal. It doesn't get any better than this. You've got the system tuned so well that you are one with the system, feeding page after page of near-perfect recognition. This is not for those who are shy, and it's not for the faint of heart. SR Realtime is possible, and it is done, but achieving it requires confidence, diligence and time. Realtime also requires the proper environment in order for it to work. I would recommend realtime for the following users:

- ✓ Experienced freelance and official reporters who are familiar with both Windows and speech recognition.
- ✓ The terminally insane

Realtime proficiency not only takes time to achieve, but it takes creativity and diligence on the part of the reporter as well. As users develop their realtime skills and develop their user files, they will discover that there are many issues that have to be dealt with, such as speaker identification, homonyms and word boundaries. Because there is no preexisting voice writer knowledge base to draw from regarding these issues, the user has to find ways to adapt their method of dictation to accommodate these issues. Certainly there is a group of users to communicate with, but much must be worked out on an individual basis, because there are many unique situations that are not shared among all users.

Speech recognition technology is evolving at an exponential rate, and software updates are numerous. Sometimes changes meant to enhance ease of use mean a whole new set of problems for realtime users, who unfortunately find out only after something has gone wrong.

Reporters also have to be realistic about their expectations regarding this new technology, and learn to recognize situations where realtime speech recognition may not be successful or is perhaps even not wanted.

Reporters must not only understand SR realtime's abilities and limitations, but they must educate those around them as well. Much of the general public is not aware or does not understand even what this technology is capable of offering, and the role of educator and liaison for this new technology falls upon the reporter.

THE HARDWARE

The hardware requirements vary according to the needs of the user. There is a wide variety of components available within each category of systems, therefore it is not possible to cover them all for purposes of this guide. Please understand that even the newest systems can encounter problems, and meeting the hardware requirements stated below does not guarantee a user will not encounter some problems. The following recommendations are based upon personal experience and reviews and recommendations from established sources in the computer industry. This is only intended to be used as a guide.

DICTATION SYSTEM

I initially began creating transcripts by redictation using an early version of Dragon Naturally Speaking in 1998, using a PII-300 with 128 MB of RAM and a SCSI hard drive subsystem. Once I put in a good sound card, I got satisfactory results.

[PC Magazine](#) December 1, 1999 issue recently evaluated all the new speech recognition products on the market using a 300-MHz Pentium II with 128 MB of RAM, running Windows 98 and a Creative Labs Sound Blaster Live! sound card and achieved excellent results. Systems that meet these specs can be bought for a very reasonable price, although you may have to upgrade the RAM and the sound card. Because I have no personal experience using lesser systems, I do not really know how they will perform. There are a great number of variables that will affect systems, too many to mention here.

TRANSCRIPTION SYSTEM

I am hesitant to make recommendations for a lesser system than one for dictation because I have not used anything less. I have read plenty of user posts of people having varying degrees of success with less powerful systems, however. If your system meets the requirements stated by the SR software manufacturer, it should be okay at least for use as a transcription system, though it may be slow in both training and its responses. If you don't use USB, you might want to make sure your sound card is one recommended for use with the SR manufacturer's

software, and preferably one of the best models you can get. For transcription purposes, audio quality, more than processor speed, is going to affect the quality results you receive. Laptop users should also check to make sure their system has been tested by the SR manufacturer.

Besides computers, there are currently only a few devices being used by most reporters for digital recording. This is because most digital recorders are not capable of economically and practically recording wave format audio files for periods of time long enough to satisfy most reporter's needs. I expect this area to change soon, however. Don't be surprised if high-capacity digital recorders of incredibly small size are available this year.

There are a number of portable devices which can be used to make a digital audio file for subsequent software transcription, and the number of options available is growing daily. This area is changing so rapidly, however, obsolescence is a concern, and many are waiting for standards to develop in this area. This boom in hand-held audio recording and playback devices is going to undoubtedly give us incredibly lightweight mobile equipment in the near future. The recorders in this area are going to be using a variety of different storage mediums, including

- flash memory cards
- compact hard disks
- digital audio tape (DAT)
- minidisk (small audio CD's)
- CDR and CDR-W
- DVD RAM (writable DVD's)

There are currently available a number of devices that are capable of recording up to one-and-a-half hours of audio in a format that can be used for transcription. The problem right now is file transfer and media costs.

There are commercial grade digital audio recorders that are being used to replace/supplement tape recorders, but they cost more than many computer systems. This high cost scares away most buyers, especially since a computer is still needed to use the audio files created by them. Macom is one of only two high-end digital recorder manufacturer that I am aware of, and they are distributed in the United States exclusively through AudioScribe. I suspect other companies will begin producing these in the future, giving us more options.

REALTIME SYSTEM

Here's where we leave PC Magazine in the dust and join the ranks of the lunatic fringe. No hardware or software evaluation to date that I am aware

of has specified recognition accuracy at a set rate of speech, nor has system performance been evaluated using continuous speech input at rates as high as 180 to 225 words per minute. Remember what I said earlier about a PII-450? I don't even know what speed systems have to run at to use Dragon Systems BestMatch III for realtime because, despite my best efforts, I haven't been able to persuade anyone to even loan me a newer system to try. So if anyone out there is willing to give a new system to me to find out for them, I'd be happy to oblige. I may be a little reluctant to give it back, however.

I will say that if you buy an 800 MHz system with 256 to 500 MB RAM, and it's still falling behind using BestMatch III, you will still get stellar results using BestMatch II, which is also still included in the software as an option. As a matter of fact, Jim Bouck and I both were using Dragon Naturally Speaking v. 3.52 at the second RVR certification testing in Athens, GA in November 1999, when Jim passed certification and I got a measly 94.63%. What I'm saying is, if you buy an 800 Mhz system, don't worry about it, it'll work for realtime.

Until recently, the only systems that could produce realtime were desktop and custom built luggable systems. The luggable, or lunchbox, systems are considerably larger and heavier than laptop systems, and they require you to be near an outlet because they require 120 volts to run. There was really no option, and actually still isn't, if you want the fastest system, because laptops lag a few generations behind desktop systems. The fastest luggable systems are now running dual PIII-800 MHz processors, which is an incredible amount of computing power in a small box. He who dies with the best toys wins.

Mobility is on its way. In late 1999, laptops became much more powerful, and now 700 MHz laptop systems, with system buses as fast as current desktops, are right around the corner. Currently, the fastest laptop you can buy runs at 500 MHz, which is more powerful than the luggable systems of less than six months ago. The only notable edge older luggables might still have over laptops is possibly faster hard drives, but even hard drives have become considerably faster over the past year, so the difference may not be noticed for most purposes. Properly configured, and with a proficient user, these laptops should be able to sustain realtime translation at speeds in excess of 200 words per minute.

THE SOFTWARE

Every bit as important as hardware is the software. If the hardware is the body of the computer, the software is the brain. Without software, computers are nothing but expensive paperweights.

SPEECH RECOGNITION SOFTWARE

An engine is a machine in which power is applied to do work by the conversion of various forms of energy into mechanical force and motion. Because speech recognition software converts various forms of information into the form of printed output, it earns the distinction of being called an engine. The speech recognition software is the engine that runs the CAT system.

There are currently four major players in the speech recognition market, but so far only [Dragon Naturally Speaking](#) has been successfully utilized for court reporting purposes. [IBM Via Voice](#) has recently made some very impressive strides in recognition accuracy, and actually achieved higher recognition accuracy than Dragon, but we do not have any hard results on its ability to translate at higher rates of speech. There are currently some court reporter users using it for redictation, but I know of none using it for realtime.

Dragon Naturally Speaking (DNS) comes in a variety of configurations, with the Preferred and Professional being the two versions most reporters purchase. The Preferred version is being purchased by those who are initially trying speech recognition. DNS Professional is the high-end version of DNS, offering features that no other version has. Of most significance to reporters is that the Pro version allows the use of multiple users and vocabularies, which comes to great advantage as a user develops his skills with the software and tailors vocabularies for specific purposes. For further information on DNS Professional features, go to [Dragon Systems](#) Web site. Also, please note that if you intend to buy a CAT system, because of the way DNS Professional is marketed, you should purchase Dragon Naturally Speaking Professional through your CAT system vendor.

CAT

Computer Aided Transcription software has been available to stenotype reporters for a number of years, and the large number of vendors and product diversity reflects CAT maturity. CAT for speech recognition is relatively new and the products are constantly changing to adapt to the incredibly rapid change that's occurring in this industry. CAT software can be bought separately or bundled with a system designed to run it. CAT is expensive to purchase, but your initial investment can be recouped in savings of not only time but in reduced typing costs.

System features range widely, from systems that are audio recorders on a PC to full-blown CAT systems, that can give formatted realtime feeds and save multi-track audio for not only playback but retranscription. Some systems are also capable of case management, which is an invaluable tool to high-volume reporters and court officials, who cover a multitude of cases on a daily basis.

Editing CAT documents has somehow earned the name scoping, probably because of the unique set of tools available to transcriptionists/scopists that are unique to this software. Again, features vary, from basic word processing tools to colloquy macros, word swap, word expansion, and numerous hotkey toggles that aid greatly in proofing a speech recognition document.

Currently, the only existing vendors in the speech recognition arena are [AudioScribe](#) and [StenoScribe](#). The [NVRA](#) Web site has a forum where the issues and features can be discussed with not only other users but also the vendors themselves. You, of course, can always contact the vendors directly with any questions you may have.

AN AUDIO QUANDARY

There's another change coming about in speech recognition technology that is going to make it more accessible to more people, and it may come as a surprise to some, because there is an ocean of change going on in this area, and one wave of one change can seem to be washed away with the tide of the next.

Most of us already know that PC processor speeds have nearly doubled in less than one year, and that hardware prices continue to drop. And by now we realize that speech recognition software companies are soaking up every added computer clock cycle like a mosquito in a blood bank, giving us higher recognition accuracy with less training time. Those are all factors that definitely have made possible the technological whirlwind that's transforming our profession before our very eyes. Right now, however, I am going to focus on what I perceive to be one of the biggest hurdles end users of speech recognition have had: audio input devices.

Sound quality has always been a key factor affecting speech recognition systems. If input audio quality isn't good, the result is poor speech recognition. While audio and acoustic engineers have been able to produce excellent recordings consistently for some time now, it has remained a challenge to the average PC user. The biggest apparent downfall has been in dealing with PC sound cards and their shortcomings.

Input audio quality with sound cards has always been a hit-and-miss proposition. Quality could be minimal to excellent, depending not only on the sound card you have but how the card behaves within your particular system. In addition, it has not been uncommon, even in newer systems, for there to be resource conflicts with other hardware within the system when upgrading these cards. When this happens, the unwary user learns the hard, slow and painful way about hardware troubleshooting and its accompanying DMA, IRQ and I/O settings. For many, this experience often results in frustration and the disgruntled user often simply giving up. I'm certain that many of you have heard war stories about those acronym demons, and that alone was probably enough to scare you away from even trying speech recognition programs.

That's the beauty of USB devices, including USB microphones: you just plug them in and they work immediately, without the hassles of setup and compatibility. Or so that's the premise anyway.

Let me briefly explain what USB is. Technically, the IEEE 1394 Serial Bus Protocol, the USB, "...targets the convergence of consumer electronics and high-speed computer peripherals." For those of us who don't speak in bits and bytes, USB means plug & play (PnP) compatibility for a wide range of computer-related devices, from keyboards and mice to microphones and speakers. It promises to offer true PnP convenience. A properly working USB device doesn't even need to have any software installed to make it work; you just plug it in and it works. No muss, no fuss, no more scrubbing. It takes personal computing into the component-based era, where you can just buy a bunch of components and plug them into a PC box and they all work together. This simplicity should finally make the PC a common household appliance.

In a typical USB mike configuration, a microphone is either connected or attaches via a plug into a USB "pod" or box that contains all the necessary components to digitally convert the audio signal and run it through the USB bus. Think of it as a plug-in external sound card and microphone. And when you think of it that way, it is easy to see that, because the sound card goes along with the mike, you can now migrate from PC to PC without having to retrain your user files. Just the concept of being able to easily transport work from home to office makes USB seem extremely attractive to me.

A problem with these USB microphones is that they are first generation products, and many will undoubtedly have at least some bugs in them. As recently as August of 1999 I was looking hard to find USB microphones. Now it seems like the market is flooded with them. How well the speech recognition software vendors will keep up their product certification with this flood of new products remains to be seen, because not too long ago there were probably less than half the products that are out now.

Also, not all computers support USB if they're older, and if a PC system is a couple years old and has USB ports, the software drivers for the USB may have problems and need upgrading. In fact, some PC systems' original USB ports may never work correctly because they were manufactured before clear standards were set within the industry. There are add-in cards on the market that can be added to a system that is either having problems with its existing USB or does not have USB, but that means cracking the case, installing a card, and possibly dealing with hardware issues, the very issue that USB was supposed to eliminate in the first place.

If you find yourself in this situation, don't rule out sound cards as a perfectly viable option at this stage of the game, either. I'm still using a sound card myself and am getting excellent results. As a matter of fact, I'm still not convinced I can get higher quality than I already have with a

USB product. And until Windows 2000 comes out, I don't intend to find out.

If you decide to upgrade your sound card, you might want to first become acquainted with what is required before actually attempting to change your sound card. To get acquainted or refamiliarize yourself with hardware, there are many resources on the internet that are free. You may want to check out [Tom's Hardware Guide](#), [Hardware Central](#) or [The PC Guide](#) for lots of great hardware info.

Once you've familiarized yourself with hardware, [Dragon Systems, Inc.](#) has a list of sound cards and devices that they have approved for use with their software on their Web site, and [IBM](#) does as well. If you have a sound card that has been tested and approved for your software, you should at least be getting usable input signal quality. If you don't, you might have some problems.

Sound cards are not all created equal, and some were, and still are, better for speech input than others. Here is a list of the best performing sound cards for speech recognition that are still available:

- Creative Ensoniq PCI,
- Turtle Beach Montego,
- Turtle Beach Fiji,
- Sound Blaster 64D Gold,
- Sound Blaster 128,
- Sound Blaster 256 Live.

So how is a person to make a purchasing decision? The best advice I can give to an upgrader is to use only USB and sound card products that have been tested by the speech recognition software manufacturers. To further your chance of success, try doing further research by reading the user support forums, watching for both positive and negative feedback. If you haven't bought speech recognition software yet, you also have the option of buying the software with USB hardware already bundled in it. Most people I have spoken to are getting at least satisfactory results using the mikes that bundled with the software, so that's a good place to start.

I have read posts from user groups across the Net that have been both pro and con for USB. The pro side says that not only is the device user friendly and convenient, but it offers superior audio quality. The con side says that USB can be buggy and that sound quality is often inferior. If you are already getting good results using your sound card, you may want to consider waiting a while before considering changing over. As for me? I am not using a USB microphone for my work system right now because I

am getting excellent sound quality using my sound card. I'm happy, and am unwilling to change right now because I don't want to fix something that isn't broken.

TO DICTATE OR REDICTATE, THAT IS THE QUESTION

In an ideal world we would only have to dictate a record while at a deposition and have a perfect transcript. In reality, however, things aren't quite so perfect. Parties talk over one another; people talk fast, mumble and have accents. You may not always have the time you'd like to prepare your system before beginning a deposition or a hearing. In a nutshell, numerous variables take an ideal situation, twist it, mangle it and contort it into something that I call a court reporter's reality. In this article, I'm going to focus on some technical aspects of dealing with these situations. I'll address the issues of controlling the environment in another issue.

With a controlled environment and parties who understand and respect the reporter's abilities when using speech recognition technology, an experienced reporter, using a developed and trained speech recognition system, can produce a record that will only need minor editing.

There is a point, however, if the system vocabulary hasn't been properly prepared beforehand, or if the dictation speed was simply too great for good recognition (consistently over 225 wpm, with high syllabic content), that you would probably be better off redictating rather than correcting a transcript.

These two situations can be mitigated somewhat, however, by a couple methods:

- 1) make an audio wave file recording and transcribe it later, or
- 2) make realtime record and an audio wave file simultaneously.

While it seems clear that the obvious choice would be to do both, most systems cannot handle both tasks and get good recognition. In fact, many systems may not be able to even make a realtime record. This would be the case with a laptop computer especially, because even the fastest models lag quite a way behind state-of-the-art desktop models at the current time.

A note for the cost conscious: You don't need a high-end realtime workstation to make an audio file. A decent laptop, properly configured, can make an audio file that can later be automatically transcribed by either that laptop or a faster computer at your leisure. The laptop option also

gives greater ease of portability and lower cost. The disadvantage is it simply is not a realtime capable system. It is a good system for a person wanting to start using speech recognition technology for transcript preparation to begin with, however.

There is another advantage to recording a digital audio file, and that is that you can actually utilize the file to train your speech recognition system. Let's assume you have dictated clearly and made a good audio file. By prompting your software to transcribe and watching it as it goes along, you can determine whether or not you're getting the level of recognition for the quality recording you have made. If you think it isn't of the quality you desire, you can stop the process and use the correction mechanism in either your speech recognition software or, if you have the ability, your CAT software. After you have made all your corrections, save your user files and transcribe the audio file again. You will most likely see a noticeable difference. Sometimes, depending on the situation, you can get even higher rates of recognition by repeating the transcribe-and-correction mechanism process two or even three times.

It is important to understand that DNS learns by the audio that is connected to the text. The acoustic information that DNS uses to recognize your voice is attached to the text in the dictation window, so if the text is gone, so is the data. When training your system, you can capitalize on this concept. Let's say you have transcribed ten pages of text in DNS and had good recognition up to the point when a faster-speaking attorney takes over, and then you notice a drop off in recognition accuracy. You probably won't want to bother spending a lot of time training your system on that first ten pages, since they are already pretty clean. You will instead want to better use your time and begin training where the speed demon took you for a ride and your recognition accuracy dropped off, which in this case was after the first ten pages.

To do this, you should delete the first ten pages from the DNS text window and not the rest, then correct using the corrections dialog, and save your user file. By doing this you will not have any errors saved from the first ten pages that you deleted because you have deleted the text. Do not delete the text you intend to correct, though, because even if you undelete it, there is no audio attached after it is restored.

Taking the above example further, if you allowed DNS to transcribe that same file to completion and/or have many pages with low recognition after that first ten pages we talked about, you only want to spend no more than 15 to 30 minutes correcting and training. If you still have uncorrected text left after you've reached that point in time, delete it before saving your user file. Here I will emphasize that making too many changes at any one time can cause errors that will result in your losing all the time you've spent

correcting. Nothing is more frustrating than to have lost an hour or more of your valuable time to a computer error.

To make myself clear, let me put it another way. If section (A) is your first ten pages, section (B) is your second ten pages, and section (C) is everything after that, then you would delete both sections (A) and (C), leaving nothing but section (B) to correct and train with on screen. Do make sure that whatever portion you choose to use to correct and train DNS has audio attached to it, though; DNS will only retain audio information with text for a certain amount of pages, determined by the settings on your particular system. Also note that I am using the number of pages loosely. In reality, depending upon how far along you are training your system, you may only get a few pages corrected in DNS in about 15 minutes' time. That's okay. In the beginning, you're safer to use your time as a gauge rather than amount of text corrected in DNS. You can always train again after you retranscribe.

It is very important to understand and remember that you do not want to save a user file if you have errors on the screen. You want it as perfect as possible. Any mistranslated words on screen that are attached to audio at the time you save your user file is like telling DNS those mistranslates are correct, and transcription accuracy will degrade as a result. For this purpose, I define an error as the software mistranslating a word you clearly said. If you stammer or mispronounce a word, I think it best to delete the word(s).

Once you have transcribed and corrected a number of times, you will notice a point where the software has reached an end result and transcription output is not improving. You should then listen to those portions of the record where it still has problems and try to determine what can be done to correct the problems next time.

One last thing: You might want to save the transcribed documents before training DNS. I have on more than one occasion deleted perfectly good text that I could have used while "battling the dragon." I am sure you will come to know, and understand, what I mean if you have decided to take the plunge into this uncharted territory, when you finally stand up from a long training session, with your eyes glazed over and your brain numb, and realize that, for all that work... you forgot to actually prepare a transcript.